Probability

1. Probabilities from Tables of Counts

The table below contains information for sales of 4 bedroom houses in Auckland City in June, July and August, 2004.

Exercise 1

	Days on the market			
Selling price	Less than 30 days	30 – 90 days	More than 90 days	**Total**
Under $300,000	39	31	15	85
$300,000 - $600,000	35	45	4	84
Over $600,000	8	4	0	12
Total	82	80	19	181

Let C be the event that a house sold for under $300,000,
D be the event that a house sold for between $300,000 and $600,000 (inclusive),
E be the event that a house sold for over $600,000,
F be the event that the sale was made in less than 30 days,
G be the event that the sale was made in 30 to 90 days (inclusive), and
H be the event that the sale was made in more than 90 days.

1. For a sale selected at random from these 181 sales, find the probability that the sale was:

(a) under $300,000 ______________________

(b) made in less than 30 days ______________________

(c) under $300,000 and made in less than 30 days ______________________

(d) under $300,000 or made in less than 30 days ______________________

____

(e) $300,000 or more ______________________

____

2. For a sale selected at random from these 181 sales, find the probability that the sale was:

(a) over $600,000 ______________________

(b) made in 30 to 90 days (inclusive) ______________________

(c) over $600,000 and made in 30 to 90 days (inclusive) ______________________

(d) over $600,000 or made in 30 to 90 days (inclusive) ______________________

(e) over $600,000 and made in more than 90 days ______________________

(f) over $600,000 or made in more than 90 days ______________________

(g) made in at least 30 days ______________________

3. The following is a student's answer to Question 1(d).

$$\begin{aligned} P(\text{C or F}) &= \frac{85+82}{181} \\ &= \frac{167}{181} \\ &= 0.9227 \text{ (to 4 d.p.)} \end{aligned}$$

Describe why this answer is not correct.

____

____

____

4. (a) Consider the event C or G (That is, house sold for under $300,000 or the sale was made in 30 to 90 days (inclusive).

Fill in the gaps: $P(\text{C or G}) = \frac{85+80-31}{181}$

$$= \frac{85}{181} + \frac{80}{181} - \frac{31}{181}$$

$$= P(\quad) + P(\quad) - P(\qquad\qquad)$$

(b) Consider the event D or H (That is, house sold for between \$300,000 and \$600,000 inclusive or the sale was made in more than 90 days).

Fill in the gaps: P(D or H) = $\frac{84+19-4}{181}$

$= \frac{84}{181} + \frac{19}{181} - \frac{4}{181}$

= P(   ) + P(   ) – P(   )

(c) Fill in the gaps: For any two events A and B

P(A or B) = ________ + ________ – _____________

5. Look at your answer to Question 2(e).

(a) Comment on the events E and H.

____

____

(b) E and H are called **mutually exclusive** events or **disjoint** events.

Fill the gap: For any two mutually exclusive events A and B

P(A and B) = __________

Discussion Exercise

	Days on the market			
Selling price	Less than 30 days	30 – 90 days	More than 90 days	**Total**
Under \$300,000	39	31	15	85
\$300,000 - \$600,000	35	45	4	84
Over \$600,000	8	4	0	12
Total	82	80	19	181

For the 181 house sales in the above table, what proportion of the houses that sold for over \$600,000 were on the market for less than 30 days?

____

____

____

2. Conditional Probability

Exercise 2

	Days on the market			
Selling price	Less than 30 days	30 – 90 days	More than 90 days	**Total**
Under $300,000	39	31	15	85
$300,000 - $600,000	35	45	4	84
Over $600,000	8	4	0	12
Total	82	80	19	181

1. For the 181 house sales in the above table:

   (a) What proportion of the houses that sold for under $300,000 were on the market for less than 30 days?

   ____

   (b) Given that a house sold for between $300,000 and $600,000 (inclusive), what proportion were on the market for more than 90 days?

   ____

   (c) If a house was on the market for less than 30 days, what proportion sold for over $600,000?

   ____

   (d) Of those on the market for more than 90 days, what proportion sold for under $300,000?

   ____

2. For the 181 house sales in the above table:

   (a) What proportion of the houses took more than 90 days to sell?

   ____

   (b) What proportion sold for under $300,000 given that they were on the market for 30 to 90 days (inclusive)?

   ____

   (c) What proportion took less than 30 days to sell or sold for between $300,000 and $600,000 (inclusive)?

   ____

   (d) Of those that sold for over $600,000, what proportion were on the market for more than 90 days?

   ____

   (e) What proportion sold for between $300,000 and $600,000 (inclusive) and took from 30 to 90 days (inclusive) to sell?

   ____

   (f) What proportion of the houses took 90 days or less to sell?

   ____

   (g) What proportion of the houses that were on the market for more than 90 days sold for over $600,000?

   ____

   (h) What proportion sold for over $600,000 and took more than 90 days to sell?

   ____

3. Comment on your answers to Questions 2(d) and 2(g).

   ____

   ____

4. (a) Fill in the gaps:

Probability a house sold for under $300,000 given that it sold in less than 30 days

$= P(\quad | \quad)$

$= \frac{\quad}{\quad}$

$= \frac{\quad / 181}{\quad / 181}$

$= \frac{P(\quad\quad)}{P(\quad)}$

(b) Fill in the gaps:

Probability a house sold for between $300,000 and $600,000 (inclusive) given that it was on the market for between 30 and 90 days (inclusive)

$= P(\quad | \quad)$

$= \frac{\quad}{\quad}$

$= \frac{\quad / 181}{\quad / 181}$

$= \frac{P(\quad\quad)}{P(\quad)}$

(c) Fill in the gaps: For any two events A and B,

$$P(A \mid B) = \frac{P(\quad\quad)}{P(\quad)}$$

3. Risk

Exercise 3

Study 1

In 1988 the results of the Physicians' Health Study Research Group study were reported in the *New England Journal of Medicine*. In this study 22 071 male physicians (aged from 40 to 84) were randomly assigned to two groups. One group took an aspirin every second day and the other group took a placebo, a pill with no active ingredient which looked just like an aspirin. The participants did not know whether they were taking aspirin or the placebo.

After five years the number of participants in each group who had had a heart attack was recorded. The results are shown in the table below.

Treatment	Heart attack	No heart attack	**Total**
Aspirin	104	10 933	11 037
Placebo	189	10 845	11 034
Total	293	21 778	22 071

1. All parts of this question apply to Study 1.

   (a) For those in the aspirin group:

   (i) The proportion who had a heart attack = ______________________

   = ______________________

   (ii) The probability that a randomly selected participant had a heart attack

   = ______________________

   (iii) The percentage who had a heart attack = ______________________

   (iv) The risk of having a heart attack = ______________________

   (v) Write this risk as a rate per 1000 participants ______________________

   (vi) Write this risk as a rate per 10 000 participants ______________________

(b) (i) For those in the placebo group, the risk of having a heart attack

= ______________________________

= ______________________________

(ii) Write this risk as a rate per 1000 participants ______________________________

(iii) Write this risk as a rate per 10 000 participants ______________________________

(c) (i) Calculate the relative risk of having a heart attack using the risk for the placebo group as the denominator (i.e., as the baseline risk).

Relative risk = ______________________________

= ______________________________

(ii) Interpret this relative risk.

For male physicians aged 40 to 84, ______________________________

____

____

(iii) Calculate the relative risk of having a heart attack using the risk for the aspirin group as the baseline risk.

Relative risk = ______________________________

= ______________________________

(iv) Interpret this relative risk.

For male physicians aged 40 to 84, ______________________________

____

____

(v) Which group is more appropriate as the baseline group? Briefly explain.

____

____

____

(d) (i) Do male physicians aged 40 to 84 who take aspirin every second day have an **increased** or **decreased** risk of having a heart attack compared to those who take a placebo?

Circle one: **increased** **decreased**

(ii) Calculate the percentage change in risk relative to the baseline (placebo group) risk.

____

____

____

(iii) Interpret this percentage change in risk.

For male physicians aged 40 to 84 ____________________________

____

____

Study 2

In 2006 the results of a study carried out among 132 271 Jewish children born in Israel during 6 consecutive years in the 1980s were published in the *Archives of General Psychiatry*. The objective of the study was to examine the relationship between father's age at birth of child (offspring) and their risk of autism.

The offspring were assessed for autism at age 17 years. The results are shown in the table below.

Father's Age Group	Autism	No autism	**Total**
15 – 29	34	60 654	60 688
30 – 39	62	67 211	67 273
≥ 40	14	4 296	4 310
Total	110	132 161	132 271

2. All parts of this question apply to Study 2.

(a) For offspring from fathers aged 15 to 29 at the birth of their child:

(i) The proportion who had autism = ______________________

= ______________________

(ii) The probability that a randomly selected offspring had autism

= ______________________

(iii) The percentage who had autism = ______________________

(iv) The risk of having autism = ______________________

(v) Write this risk as a rate per 10 000 offspring ______________________

(b) (i) For offspring from fathers aged 30 to 39 at the birth of their child, the risk of having autism

= ______________________

= ______________________

(ii) Write this risk as a rate per 10 000 offspring ______________________

(c) (i) For offspring from fathers aged 40 or more at the birth of their child, the risk of having autism

= ______________________________

= ______________________________

(ii) Write this risk as a rate per 10 000 offspring ______________________________

(d) Using the risk for fathers in the 15 – 29 year age group as the baseline:

(i) Calculate the relative risk for fathers in the 30 – 39 year age group of having autistic offspring.

Relative risk = ______________________________

= ______________________________

(ii) Interpret this relative risk.

____

____

____

(iii) Calculate the relative risk for fathers in the 40 or more year age group of having autistic offspring.

Relative risk = ______________________________

= ______________________________

(iv) Interpret this relative risk.

____

____

____

In parts (e) and (f), use the 15 – 29 year age group as the baseline group.

(e) (i) Do fathers aged 30 to 39 have an **increased** or **decreased** risk of having autistic offspring compared to those aged 15 to 29?

Circle one: **increased** **decreased**

(ii) Calculate the percentage change in risk relative to the baseline risk.

(iii) Interpret this percentage change in risk.

(f) (i) Do fathers aged 40 or more have an **increased** or **decreased** risk of having autistic offspring compared to those aged 15 to 29?

Circle one: **increased** **decreased**

(ii) Calculate the percentage change in risk relative to the baseline risk.

(iii) Interpret this percentage change in risk.

4. Independent Events

If A and B are independent events then P(A | B) = _______________

Exercise 4

Selling price	Days on the market: Less than 30 days	30 – 90 days	More than 90 days	**Total**
Under $300,000	39	31	15	85
$300,000 - $600,000	35	45	4	84
Over $600,000	8	4	0	12
Total	82	80	19	181

1. Recall: C is the event that a house sold for under $300,000 and
   F is the event that the sale was made in less than 30 days.
   Are events C and F independent?

   ____

   ____

   ____

2. Recall: E is the event that a house sold for over $600,000 and
   H is the event that the sale was made in more than 90 days.
   Are events E and H independent?

   ____

   ____

   ____

Discussion Exercise

Fill the gaps: For events A and B, $P(A\,|\,B) = \dfrac{P(\qquad\qquad)}{P(\qquad)}$

If A and B are independent events, $P(A\,|\,B) = P(\qquad)$

So if A and B are independent events, $\dfrac{P(\qquad\qquad)}{P(\qquad)} = P(\qquad)$

That is, if A and B are independent events $P(\qquad\qquad) = P(\qquad) \text{ x } P(\qquad)$

Exercise 5

1. For events A and B, P(A) = 0.8, P(B) = 0.5 and P(A or B) = 0.9.

   (a) Calculate $P(A \cap B)$

   (b) Calculate P(A | B)

   (c) Are A and B independent events? Justify your answer.

2. For events A and B, P(A) = 0.3, P(B) = 0.4 and P(A | B) = 0.

   (a) Are A and B independent events? Justify your answer.

   (b) Are A and B mutually exclusive events? Justify your answer.

3. Events A and B are independent. Also P(A | B) = 0.4 and P(B) = 0.6.

   (a) State P(A)

   (b) Calculate P(A and B)

4. Events A and B are mutually exclusive. Also P(A) = 0.4 and P(B) = 0.3.

(a) State or calculate P(A | B)

(b) Calculate P(A or B)

5. Tables of Counts and Probability Trees

Blood Group Systems Example: Example 4.7.3 from Chance Encounters (p 179)

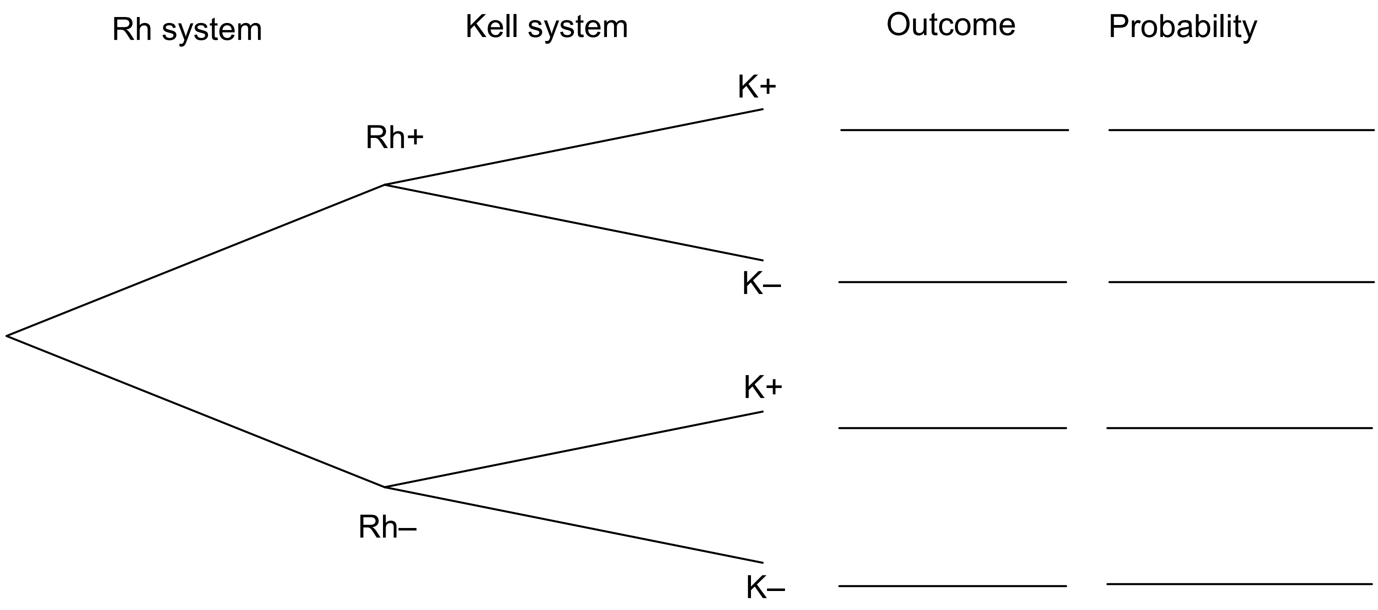
There are a large number of genetically based blood group systems that have been used for typing blood. Two of these are the Rh system (with blood types Rh+ and Rh–) and the Kell system (with blood types K+ and K–). It is found that any person's blood type in any one system is independent of his or her blood type in any other.

It is known that, for Europeans in New Zealand, about 81% are Rh+ and about 8% are K+. If a European New Zealander is chosen at random, what is the probability that he or she is either positive in both systems or negative in both systems?

Table of Counts Approach

	Kell system		
Rh system	K+	K–	**Total**
Rh+			
Rh–			
Total			

Probability Tree Approach



Imperfect Testing Procedures Example:

The ELISA test is used as a screening test for HIV. The test, however, is **not** a perfect one.

For people with HIV: 99.7% test positive

For people without HIV: 0.3% test positive

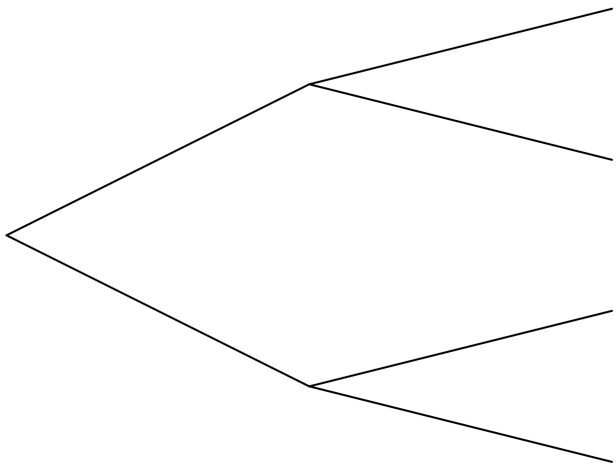
It is estimated that 0.1% of the New Zealand population have HIV.

Question: Suppose that a person is picked at random from New Zealand.
What is the probability of having HIV given that the test is positive?

Table of Counts Approach

	Test result		
Disease status	Positive	Negative	**Total**
HIV			
Not HIV			
Total			

Probability Tree Approach



Outcome | Probability